AD-A258 367

COMPUTER ASSISTED MEDICAL DIAGNOSIS PROBLEMS AND METHODS TO MINIMIZE THEIR EFFECTS

D. H. Ryman

Report No. 91-32

Approved for public release: distribution unlimited.

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
SAN DIEGO, CALIFORNIA 92186-5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

COMPUTER ASSISTED MEDICAL DIAGNOSIS PROBLEMS,

AND METHODS TO MINIMIZE THEIR EFFECTS

DAVID H. RYMAN

Naval Health Research Center

P.O. Box 85122

San Diego, CA 92186-5122

Report 91-32, supported by the Naval Research and Development Command, Department of the Navy, under Work Unit No. M0095.005-6051. The views expressed in this article are those of the author and do not reflect the official policy or position of the Department of Defense, nor the U.S. government. Approved for public release, distribution unlimited. An earlier version of this paper was presented at The 5th Annual Conference for Learning Technology in Health Care Sciences, Orlando Fl. March 21-23, 1991.

SUMMARY

Objective

The background and development of Computer Assisted Medical Diagnostics (CAMD) systems were reviewed and the problems and errors encountered during the develop-ment of such systems were examined. Five specific methods for projecting a diagnosis--Expert Rule Based, Bayesian, Statistical, Neural Networks, and Decision Trees--are discussed in terms of the problems previously reported in the literature and the solutions to those problems, or ways to minimize the problems.

Approach

A review of the literature indicates multiple experts should be used when developing rule based systems. Requiring adequate sample sizes and obtaining multicenter samples with a sufficient number of representative disease cases has been recommended in the development of Bayesian systems. Statistical systems have been criticized because they do not offer users a clear explanation or interpretation, while Decision Trees can become too complex when they become large. Often, system limitations can be acertained by conducting a thorough validation procedure which tests the system using specific disease cases, identifies signs and symptoms that were omitted during system development, and reveals diagnoses that have similar presentng problems signs and symptoms which are not addressed by the system.

The utilization of sign and symptom disease complexes and the use of multiple disease diagnostic methods to identify possible disease diagnosis is emphasized. Furthermore, it is recommended that the level of diagnostic expertise of the user should be considered, and that CAMD systems should give standard definitions and interpretations for any signs, symptoms, lab results, diseases, or treatments suggested.

Conclusion

Using a data base management approach when developing a CAMD system allows multiple methods of disease prediction to be integrated. A capability to present standard definitions and information on methods for gathering or eliciting signs and symptoms should be available. Also, a CAMD system should provide information regarding standard treatments, disease progression, and possible complications, as well as indicating the reasons behind the prediction given by the system. CAMD systems that are easy to modify, edit, and change will remain in use and form the basis for evolving systems.

COMPUTER ASSISTED MEDICAL DIAGNOSIS PROBLEMS, AND METHODS TO MINIMIZE THEIR EFFECTS

DAVID H. RYMAN

In the late fifties, Ledley and Lusted (1959) suggested that the decision process for arrivng at a medical diagnosis could be modeled using a medical combination of symbolic logic and conditional probability methods. Barnett (1982) pointed out that computer programs for medical diagnosis had "the greatest potential when the clinical problem is relatively well defined and structured and when only a limited number of diseases need to be considered." Yet, five years later Barnett et al. (1987) wrote the introductory article for DXplain, a Computer Assisted Medical Diagnosis (CAMD) system that included over 2,000 diseases and 4,700 signs and symptoms with a knowledge base that specified more 65,000 relationships among them. The development of computer systems like CASNET (Weiss, Kulikowski, Amarel, Safir, 1978), INTERNIST (Miller, Pople & Myers), CADUCEUS (Blois, 1980), and MYCIN and ONCONIN (Shortliffe, 1976) for diagnosing diseases, managing treatments, consulting or explaining has progressed rapidly, and more comprehensive disease diagnostic systems like DXplain (Barnett, et al. 1987) and INTERNIST-1/QMR (Miller, et al. 1982) have evolved. Most CAMD systems today are limited in scope to a few related diseases or only one disease. This paper discusses some sources of the problems, and limitations in CAMD systems including errors in development, lack of standardization or definition for signs and symptoms, and problems with methods or algorithms used in suggesting diagnosis. Solutions found in the CAMD literature are also reported.

General CAMD Functions and Methods

Shortliffe's review of medical decision-support systems (1987) listed three functions of such systems; a) Information management, b) Focus attention by flagging abnormal values, explaining possible abnormalities, or alerting possible drug interactions, c) Patient specific consultation and assessment. Although assisting in making a diagnosis and identifying appropriate treatments or tests are two functions that are often incorporated together,

many CAMD systems include only one of these functions.

CAMD systems obtain information either; a) actively, by monitoring medical devices or medical records to give warnings, advice, or report on conditions, or b) passively, by waiting for sign and symptom inputs. Most CAMD systems are passive data systems that either make suggestions or critique the information or decisions that are entered. The CAMD systems also differ in Human Factors areas, which include ease of use, reliability of computer system, ease of entering information (mode and length process), and the degree to which displays are informative and easily read.

General and Developmental CAMD Problems and Error Sources

In this section those development problems that apply to all types of CAMD systems and possible solutions to these problems will be discussed. All CAMD methods depend on knowledge from experts or analysis of data from large samples of cases. This information must be represented in terms of cases, and data for the problem area. If CAMD systems are developed with only well measured variables on well defined cases (e.g., selected cases at Medical Research Hospitals), and applied in more general conditions (e.g., all emergency rooms), then shrinkage in diagnostic accuracy is to be expected. The cases that are discarded during, or before system development, should be tried in the final system to estimate the CAMD 'worse case' performance. A caveat about the limitations of disease area and the cases to be used should be made explicit in any CAMD system.

All diseases included in the system must be adequately represented in terms of number of rules or data cases. The age and sex of patients, geographical areas and other important epidemiological factors must be adequately sampled. The quality and accuracy of data used in the CAMD system development should be realistic in terms of the conditions in which the final

system is to be applied. The accuracy of many signs and symptoms that are routinely taken vary widely. Vital sign data like temperature and blood pressure can be influenced by; patient conditions (warm or cold liquids taken, recumbent versus erect), method of measurement (electronic temperature and sphygmomanometers versus mercury thermometer, and cuff and stethoscope), as well as varying between the persons taking the measurements (length of time taken, hearing acuity or significant beat determination). Henderson, Moeller, Ryack, and Schumack showed that training in the collection of data, especially in judgmental clinical symptoms (e.g., severity of pain, site of pain, etc.) improved the accuracy of an Abdominal Pain Bayesian CAMD program used by U.S. Navy corpsmen from 48% to over 70% correct diagnosis. This can be done by presenting standardized term definitions and descriptions at the bottom of the screens in CAMD programs. Experts should be used to delineate these standard set of measures necessary for the CAMD diagnostic area, including standard term definitions and measurement procedures. When certain important information is omitted some CAMD systems automatically request that information. This can prevent incomplete and unequal information bias across cases, as well as provide the information necessary for a complete and standard medical history.

CAMD systems should request information in a clear and logical manner. Systems which gather information in the order in which it is usually collected enable direct computer input (i.e., past to current medical history, examination results, laboratory findings). The interpretation of the method used in arriving at diagnosis (significant signs and symptoms for suggested diagnosis) should be available to the users. Often CAMD systems use 'why', 'how', or '?' as prompts to get the CAMD system to indicate how the 'opinion' was derived. Some systems give medical literature citations for the significant sign and symptom disease indicators, while many expert systems list only the rules 'fired' or those 'significant' signs and symptoms input into the CAMD system that led to a given diagnosis.

Methods and Algorithms Used in CAMD Systems

Tuhrim and Reggia (1986) and Williams (1982) listed and described four diagnostic inference methods; a) categorical, b) probabilistic, c) artificial intelligence, and d) pattern recognition. Sneiderman (1986) lists examples of CAMD systems for each of these four inference methods. Recently Eberhart and Dobbins (1990) employed another methodology, neural networks, to make diagnostic predictions. Table 1 lists these five methods and features of the algorithms used. Rule based systems and neural networks are both from the field of artificial intelligence, but are so different that they are described here separately. Also, those systems referred to as Decision Trees have been called Categorical Systems.

Though the various methods and algorithms will be discussed separately in terms of their problems and errors, many CAMD systems combine several of these methods in suggesting diagnosis. The Bayesian Acute Abdominal Pain CAMD system used on U.S. Naval ships (Carras, Southerland & Fisherkeller, 1989), has an initial rule "If Female and Location of Pain is Flank or lower Abdominal then suggest Female Abdominal Pain Bayesian Program." Another Abdominal Pain program (Sturman & Perez, 1989) combines simple conditional rules ("If Appendectomy History then Appendicitis = 0") with a special algorithm that combined, correlated, matched, and weighted patient information. Some decision tree analytic methods produce results that look like simple rule based hierarchies, but in fact, utilize algorithms from the mathematical field of decision tree analysis that could be viewed as specialized statistical solutions.

TABLE 1

FEATURES OF DISEASE DIAGNOSIS ALGORITHMS AND METHODS

RULE BASED EXPERT SYSTEMS

- Rules using signs and symptoms
- Forward or backward solution
- Rule based disease likelihoods

BAYESIAN

- Initial *a priori* disease probabilities
- Conditional sign and symptom disease probabilities
- A *posteriori* disease likelihoods

STATISTICAL METHODS

- Sign and symptom weights
- Scales of signs and symptoms
- Regression Methods (non-linear, logistic)
- Multiple discriminant solutions

NEURAL NETWORKS

- Learning rate, momentum
- Number of iterations
- Number of hidden nodes and layers
- Node activation function

DECISION TREE LOGIC

- Root node (disease area)
- Non-terminal nodes (disease categories)
- Terminal node (specific disease)

Expert Rule Based Method - Problems and Suggested Solutions

Many commercially available Expert Shell Systems allow 'Medical Experts' to quickly develop CAMD applications that rely heavily on variations of the "if - then" rule statement to process information and suggest medical decisions or diagnosis. Other systems utilize logic trees, structured solutions (semantic nets or frames and domains), likelihood weighting, Bayesian analysis, and a few systems have the ability to generate rules from sample cases or data.

Many of these Expert Systems can 'justify' the suggested diagnosis or treatments by listing the rules used or providing comments or medical references. One physician (Kinney) has published articles in three different diagnosis areas using different expert shell systems (Kinny, Cortada, Galbut, Larsen, 1986; and Kinney, Brafman & Wright, 1988) each time finding an easier, more sophisticated shell system on which to develop a rule based CAMD application. Some of the problem areas in building a knowledge base including decisions regarding which experts to use, how many experts to use, how much information to process, and how many rules to include. The ease of use of the expert system and the necessity of using a 'knowledge engineer' or programmer, also varies between these Expert Shell systems.

Hughes, Gose, and Roseman (1990) listed four problems presented in most medical expert system;

1) The static nature of 'finished' systems (not allowing for new medical findings). Adding new rules to some Expert systems creates changes in the rule firing order progression from the initially tested version.

To prevent this, the initial rules should be well thought out by a number of experts, and thoroughly tested at each stage of development.

Systems that are more flexible in adding or modifying rules should help overcome this problem.

2) Limited to medical specialty addressed. Difficult cases can often present signs and symptoms that 'look like' a certain medical area but are not.

If the Expert system used can pass and receive information, then multiple CAMD systems can access the data from a particular case (shared data bases). Most Expert systems now accept and output standard data base formats (ASCII, DBase, Lotus 1-2-3), and this enables various CAMD systems to utilize the same case data.

3) Cannot easily use geographical medical knowledge disease information (disease differences between geographical regions).

Rewriting all the rules for each geographical region is not feasible, but incorporation of initial disease rates by region may be more reasonable. Sometimes the difference in occurrence of disease by geographical region will be reflected in the occurrence of related signs and symptoms. Zagoria & Reggia (1983) and Zoltie, Horrocks & de Dombal (1977) have found that Bayesian CAMD systems were 'portable' between geographically diverse institutions. Some expert systems shells (Henderson, Moeller, et al., 1978) allow for the use of the Bayesian method.

4) Leads to suboptimal results. Truth is seen as 'the consensus of experts' and the best recognized experts maybe too busy for the lengthy knowledge engineer interviews in construction of the necessary knowledge base rules formation.

A solution to this problem is to select medical experts to develop rules for the specific disease(s) for which they are recognized as experts.

Comparing the rules and results of test cases between such individual experts systems would show the best sets of rules for specific diseases. Medical experts in the future may be expected develop expert CAMD systems as frequently as they write medical texts today. This will occur with the increased ease of use, and with wider exposure of these systems.

Some commercially available Expert Shell systems that allow for the quick development of CAMD applications produce a minimum set of questions to arrive at a diagnosis. While this may be suitable for some CAMD applications, it may preclude them from gathering a full set of signs and symptoms necessary to gather a complete medical history so necessary in many disease areas. Many of these shell systems also include error and logic checking features, but this does not prevent medical experts checking all the final rules in any CAMD system. Another problem is the ability of these systems to handle uncertainty, 'Fuzzy Logic' (degrees of truth) and Fuzzy sets (group membership not crisply defined) differs widely between systems.

Bayesian Method - Problems and Possible Solutions

Minasi, (1990) listed four 'flaws' in the Bayesian Method.

1. Bayesian systems cannot discriminate between important and unimportant questions (the method of showing importance is not incorporated, but reflected by the sign and symptom occurrence differences between diseases).

To indicate importance, Reggia and Perrione (1985) used a combination of rank ordering patient sign and symptom disease probabilities, listings of patient signs and symptoms yielding zero disease probabilities, and listings signs and symptoms that contributed to low probability diseases to indicate importance. Habbema and Gelpke (1981) reported a Bayesian discriminant

analysis program that could select significant variables for determining a diagnosis.

Sign and symptom disease conditional) probabilities for a disease can easily be tested for significance (Analysis of Variance followed by a posteriori tests like Scheffe's or Tukey B tests) when total sample size and initial (a priori) diseases are presented, or the number of cases for each disease on which they are based is known. Such analysis will show the specific sign and symptom alternatives that are different between specific diseases. This approach will indicate which sets of signs and symptoms discriminate specific diseases for any Bayesian CAMD system. It is not necessary, however, to apply a statistical analysis to these Bayesian sign and symptom disease probabilities. Experts can easily persue the probabilities for each sign and symptom alternative across all diseases to locate the signs and symptoms factors in each disease appear medically reasonable or correct. Expert verification or selection of important sign and symptom alternatives for each disease can give credence to the suggested diagnosis. These methods can also suggest scales of signs and symptoms or indexes of symptom disease complexes.

2) Bayesian systems not know when to stop because further questions will not affect the outcome.

This is not necessarily a problem. Sometimes a standard set of questions is necessary for a particular set of diseases, and it can be argued that administration of a full standard examination should be encouraged so the complete medical case history will be gathered.

3) Fact uncertainty can not be incorporated (only yes/no responses).

One can leave out signs and symptoms with uncertainty or assume that uncertainty is reflected by conditional probabilities (sign and symptom

occurrence for each disease). Some expert systems utilizing fuzzy logic incorporate such uncertainties (Hudson & Cohen, 1987) by assigning Bayesian like probabilities. Habbema and Gelpke (1981) utilized two methods (treat as category and skip variable) in handling missing data (the ultimate uncertainty). Alternatively, the collection of sign and symptom case data to determine the conditional probabilities should include responses like unknown, possible or uncertain. When these responses are frequent enough and significantly different between diseases the conditional probabilities would be available for inclusion.

4) Forward chaining (progression from sign and symptoms to disease) is the only method allowed.

Though this is true of the classical Bayesian method, a Bayesian CAMD system where the significant conditional probabilities have been determined could present these significant signs and symptoms when 'backward chaining' from disease to expected or significant signs and symptoms is desired.

Another Bayesian method problem is that it requires a large quantity of 'good' data in development. Some investigators believe that the frequency of occurrence of particular symptoms with specific diseases are virtually impossible to obtain from even good hospital records (Edwards, 1972). In addition, that symptom and disease diagnosis data have too much error, often varying from year-to-year and doctor-to-doctor, and that symptom-disease complexes and disease categories change too much. This belief is supported by the fact that the International Classification of Diseases has changed ten times in the last twenty five years. In addition, text books and journals rarely state specific probabilities of symptom- disease occurrence or describe them in terms like 'usually', 'often', 'rarely.'

Finally, developed Bayesian systems are also seen as static, non-changing, non-evolving. Periodic updating of disease occurrence or sign and

symptom disease occurrence, when they change, would resolve this. Hybrid systems that link data bases and update Bayesian disease rates, and associated conditional sign/ symptom disease conditional probabilities will overcome this problem. The classical Bayesian method does not allow for multiple diagnosis. Ben-Bassat (1980, 1983) has modified the Bayesian method to allow for multiple diagnosis by developing multiple Bayesian solutions for each disease (solving for each disease verses all other diseases for all diseases using the Bayesian method for each disease separately). In addition, the problem of rare diseases and how many sign and symptom predictors to use for them has been addressed. Fryback's (1978) findings suggest that using a small number of good signs and symptoms disease indicators produces a better Bayesian solution for rare diseases and Charniak (1983) suggests using the log of probabilities where the diseases and associated signs and symptoms very widely.

Bayesian Assumption of Independence

Much has been written about the Bayesian reliance on assumption of independence. However, Fryback (1978) stated,

> **"the greater the number of variables used in the Bayesian calculations, the more the degradation of the model's performance when the data violate conditional independence. When relatively few variables are used, the independent contributions of each toward making the final Bayesian diagnosis seem to outweigh the degradation in performance due to over weighing redundant information. As the variable set increases in size the independent contribution gained by adding yet one more variable tends to be less tan the detrimental effect of multiply counting its information already conveyed by other variables" (p 433).**

He also suggest using a small number of the best diagnostic signs and symptoms (even if highly related) to maximize Bayesian performance. Charniak (1983) did not find independence assumptions to be a problem, stating "it is so bad that it is fortunate that it doesn't matter" (p. 71). He suggested introducing "pathological states and causal reasoning -- something most AI programs do

anyway" (p. 73). Russek, Kromal and Fisher (1983) suggested that "choosing independence and using many variables may well provide a better classification rule than keeping fewer variables and using the full model" (p. 550). Chard (1983) concluded, "diagnostic efficiency of Bayes theorem will not be greatly influenced by dependence if a reasonable amount of common sense is applied to the selection of the knowledge-base, with either elimination or correction to the most obvious dependencies" (p 19). Hilden (1984) has even derived a version of Bayes Formula ('Relaxed Model') that does not require the conditional independence assumption.

With regard to the above discussion, it is concluded that combining highly interrelated signs and symptoms that differentiate specific diseases can overcome the Bayesian problem of assumption of independence. The symptoms of cough, sputum, nasal congestion are considered Upper Respiratory Infection (URI) indicators, and as a sign and symptom disease complex are often mentally summed together. Other sign and symptom disease complexes may not be as highly correlated but are considered together as a scale, for example the Glascow Comma Scale that is widely used in Emergency Rooms, as indicative of the severity of a case. Graham (1977) reported a scale with seven signs and symptoms that differentiated acute appendicitis from gangrenous and perforating appendicitis.

Statistical Method Problems and Possible Solutions

Wittkowski (1990) in reviewing Statistical knowledge-based systems mentioned the differences/problems in these systems including:

1) User familiarity with such systems (little acceptance without clear explanation or interpretation).

2) System development difficulty because of different types of knowledge used (medical knowledge rarely states sign and symptom disease relationships

in formulas, though more and more medical can be interpreted by deterministic formulas).

3) Bad track record (many statistical based CAMD systems have failed, mainly because they were too ambitious or based on rules that were misleading enough to impair performance/decision adequacy).

Rules should be checked with other human experts.Carrol (1987) argued that simple mathematical models (linear regression could more accurately estimate underlying expert decision maker's processes than expert systems stating; "Given the degree of error that exists in the real world in both the nature of decision problem, and the measurement of predictor and criterion variables, linear models simply do a better job than other algorithms" (p. 290). Hughes, Gose, and Roseman (1990), also commented that of the three sources of medical knowledge formal instruction, medical literature, and years of patient contact statistical analysis of large patient populations dominate the medical literature. They also believe that expert opinion is also derived from experience and statistical assessment of data. In fact, relatively few CAMD systems rely solely on statistical methods like the various regression methods or discriminant functions.

Neural Network Method Problems and Possible Solutions

Neural Networks can be viewed as expert systems that process data to arrive at mathematically based 'rules' that contribute to decisions. The three basic parts to a Neural Network includes:

1) Architecture (the interconnections of the variables and nodes in the system).

2) Activation Function (the mathematical way of 'firing' nodes).

3) Learning Rule (the method which develops the variable weights which compute the nodes in the network).

The model for such systems is that of the human brain, where incoming information is processed and neurons are fired when significantly stimulated. Dunbar (1989) discussed some difficulties in developing and designing Neural Network systems including:

1) Determining of number of nodes and number of hidden layers.

2) Determining the learning rate and momentum for the given number of input and output variables, and sample size.

Neural Network programs generally require the user to specify the number of nodes, number of hidden layers, the learning rate and momentum from samples of data. Automatic determination of these parameters would be a valuable feature to incorporate into a neural network program. It would also be useful to develop an analytical method for determining these parameters.

3) Interpretation of resulting systems (weights, hidden nodes).

Caudill (Eberhardt & Dobbins, 1990) recommends that expert systems be used to provide explanation for neural network structure, node antecedents, and consequents. Some systems display graphical representations (Hinton diagrams) interpreting neural network weights and nodes (Casenet).

4) Large number of sample cases required.

Neural Network systems use a 'training' set of cases. Generally, preliminary analysis of those data are used to determine the number of nodes and layers. The final set of weights derived for these nodes using the training set.

Another set of cases, the learning set, is used to evaluate the predictive probability of those weights.

Decision Tree Method Problems and Suggested Solutions

Decision trees can visually represent the progression of diagnosis (Goldman, Weinburg, Weisburg, & Olsen, 1982) as steps in the information gathering process, indicate the important medical decision stages (Clarke, 1984), or show the sign and symptom complexes at every level of disease differentiation (Kurzynski, 1987). Clarke (1984) found two problems in a larger decision tree; a) a larger number (>15) of pieces of information at each stage was required and, b) there were unintended shifts between disease states across the tree. Dubois and Brook (1988) also recommended the use of Bayesian logic and pruning (reducing the number of nodes) of decision the trees.

Kurzynski (1987) used two decision stage mathematical techniques (nearest neighbors and modified multistage classifier) in developing a three stage classification system for diagnosing 16 Abdominal Pain diseases. He found that seven signs and symptoms were needed to maximized decisions at each stage.

Fleiss (1972) compared Bayesian, statistical (discriminant functions) and logical decision tree methods in psychiatric diagnosis and found that none stood out as superior, but that the decision tree method had three advantages (did not require large developmental sample sizes, generalized better to new populations, and could be used for categories where few cases existed).

Combination (Hybrid Systems) - Overcoming Single Method Problems

In presenting the problems of methods and algorithms used in CAMD systems, it is apparent that, while some of these problems are shared between

the various methods many are unique to a particular method or algorithm. The utilization of multiple methods and algorithms could be used to overcome the limitations of any particular CAMD system's reliance on a single method or algorithm. Schwartz, Petit and Szolovits (1987) in a general critical review of Artificial Intelligence in Medicine pointed out the "irony that early programs strategies after being discarded might enhance the performance of new programs" (p. 687). Macartney (1987) pointed out "no method of diagnosis helped by computers has been shown consistently to be superior to all others' (p. 1331). Caudill (Eberhardt & Dobbins, 1990) suggested using expert systems in parallel with neural network systems, so that expert systems could explain a network's operation by processing the network's input data and final decision using backward chaining to indicate the network's reasoning.

Berger, Gelfand, and Miller (1990) described a system to manage diabetes mellitus that utilized patient glucose values and insulin dose history in a CAMD system which used rule based logic and statistical methods along with a physiological model involving insulin and glucose to optimize insulin dosage and detect patterns and trends in glucose-insulin data. Sturman and Perez (1989) used a "special computational algorithm" that combined, correlated, matched and weighed patient information and utilized conditional rules to arrive at abdominal disease likelihoods.

The utility of the computer in managing and storing data should not be overlooked by CAMD systems. De Tore (1988) suggested that medical microcomputer applications "should not 'stand alone', but should be integrated into medical information management systems to insure medical, as well as, administrative aspects of these systems." Similarly, Greenes, et al (48) pointed out the potential of the "computer approach called 'knowledge management'" in both clinical problem solving and medical education.

Tuhrim and Reggia (1986) developed an "expert system generator" which medical students used to implement medical expert systems quickly. Huges,

Gose and Roseman reported a hybrid system combining an expert system and a statistical analysis system which could refresh the statistical correlation of each rule ensuring a dynamic, current, statistically accurate rule base (p 63). Finally, Lachman (1989) has developed a multiple expert system with forward and backward chaining along with Bayesian systems and rule updating.

Discussion

The use of such combination/hybrid systems allows for the testing of agreement behind the scenes 'in the background'. It would seem that if several methods are used as diagnostic indicators for certain diseases, and they agree, there is more certainty than single CAMD systems which use only one method or algorithm. When one CAMD method is superior for a particular disease, and that disease is suggested by that method or other equivocal methods, then one must determine which method is best. Nordyke, Kulikowski & Kulikowski (1971) compared Bayesian, statistical, and pattern recognition methods in diagnosing thyroid dysfunction states, using three areas (stages) of information (Medical History, Physical Exam and Laboratory Results). They found, "each of the methods uses the characteristics of a patient differently, some taking advantage of discriminating information at a given stage better than others, it would seem that a combination of these would be best for a sequential diagnostic procedures" (p. 389).

Gino, Pugh and Ryman (1990) reported a MUMPS language based system that uses, creates, and maintains patient historical, medical, diagnostic, and treatment information. Integrate Bayesian, Expert Rule Base, neural network, and statistical systems In the development of this system the Database Management Approach was emphasized; patient information is stored in formats that can be utilized both by the various methods and algorithms. Also, the combination of signs and symptoms by either unit weighing, regression, or neural network weighing is incorporated in this system. This is a menu driven diagnostic shell system that allows question screens to be easily constructed

with weights from any algorithm method directly associated with any response alternative. Diseases and treatment descriptions and definitions can be easily entered with menus and screens. The rule based expert system incorporated with this system was modified from the RULEMAN system developed by Dymond (1982). The Neural Network and statistical solutions are incorporated in subroutines using any weights assigned to any sign and symptom alternative. The Medical Practice Support System (MEPSS) currently under development (Stetson, Eberhart, Dobbins, Pugh, & Gino, 1990) is being designed as an extension to the system developed by Gino et al (1990).

There will be diseases in any CAMD system in which any or all methods will not produce a diagnostic accuracy higher than certain physician groups (specialists, senior medical staff). When this occurs CAMD systems should indicate this, especially when they are used by these physicians. For example, when these physician users diagnosis a disease where they are found to be more accurate than the CAMD system, then the system should only confirm if it can do so. CAMD systems, even in such situations, still can be useful in rapidly intaking the information, writing such summary reports requested, and storing permanent patient records. With the use of CAMD systems that allow for multiple algorithms and methods, more individual disease area applications can be incorporated quickly. This is especially important considering the number of these applications in existence and being developed.

Conclusion

This paper has presented some of the general and developmental difficulties in CAMD systems, and the criticisms of the methods and algorithms used by these systems in suggesting diagnoses. Suggestions for overcoming or limiting the effects of most of these problems were reviewed. The advantages of computers, in rapidly accessing information, presenting standard requests for signs and symptoms, and in storing vast amount of data accurately are just some of the obvious advantages of CAMD systems. The comparisons of CAMD

diagnostic accuracy with the various levels of medical users, and with the various methods and algorithms, should be included along with any suggested diagnosis or treatments given by the system. With the interconnections of computers in networks and over modems CAMD systems no longer have to be limited to single data sources, methods, knowledge bases, or even single systems. The flexibility of CAMD systems in terms of adding to, modifying, and updating them, will make them valuable as sources of second opinions over longer periods of time. Continued critical comments and analysis of CAMD systems and their methods should improve them and the methods used by them in suggesting diagnosis.

REFERENCES

Barnett, G.O. (1982). Editorial - The Computer and Clinical Judgement, The New England Journal of Medicine, 307, 493-494.

Barnett, G.O., Cimino, J.J., Hupp, J.A. et al, (1987). DXplain: An Evolving Diagnostic Decision-Support System, JAMA, 258, 67-74.

Ben-Bassat, M. (1980). Multimembership and Multiperspective Classification: Introduction, Applications, and a Bayesian Model, IEEE Transactions on Systems, Man and Cybernetics, SMC 10, 331-336.

Ben-Bassat, M., Campell, D.B., MacNeil, A.R. & Weil, M.H. (1983). Evaluating Multimembership Classifiers: A Methodology and Application to the MEDAS Diagnostic System, IEEE Transactions Pattern Analysis & Machine Intelligence, PAMI 5, 225-229.

Berger, M.P., Gelfand, R.A. & Miller, P.L. (1990). Combining Statistical, Rule-Based Methods to Assist in the Management of Diabetes Mellitus, Computers & Biomedical Research, 23, 346-357.

Blois, M.S. (1980). Clinical Judgement and Computers, New England Journal of Medicine, 303, 192-197.

Carol, B. (1987). Artificial Intelligence - Expert Systems for Clinical Diagnosis: Are They Worth the Effort, Behavioral Science, 32, 274-292.

Carras, B.G., Southerland, D.G. & Fisherkeller, K.D. (1989). MEDIC - ABDOMINAL PAIN; A Decision Support Program for the Management of Acute Abdominal Pain, (Naval Submarine Medical Research Laboratory Tech Report 1146), Groton, Connecticut.

Chard, T. (1983). The Effect of Dependence on the Performance of Bayes' Theorem: An Evaluation Using a Computer Simulation, Computer Methods & Programs in Biomedicine, 29, 15-19.

Charniak, E. (1983). The Bayesian Basis of Common Sense Medical Diagnosis, Proceedings of the National Conference on AI, AAAI, 70-73.

Clarke, J.R. (1984). A Concise Model for the Management of Possible Appendicitis, Medical Decision Making, 4, 331-338.

DeTore, A.W. (1988). Medical Informatics: An Introduction to Computer Technology in Medicine, The American Journal of Medicine, 85, 399-403.

Dubois, R.W. & Brook, R.H. (1988). Assessing Clinical Decision Making: Is the Ideal System Feasible?, Inquiry, 25, 59-64.

Dunbar, J. & Gino, A. (1989). Neural Networks and Their Possible Use in Computer-Assisted Diagnosis, (Naval Health Research Center Technical Report 89-42).

Dymond, A.M. (1982). An Expert System for the DHCP Environment, Mumps User's Group Quarterly, 18, 6-9.

Eberhardt, R.C. & Dobbins, R.W. (1990). Neural Network Versus Bayesian Diagnosis of Appendicitis, Proceedings IEEE EMBS Annual Conference, Philadelphia, PA, Nov. 17-19, 78-80.

Edwards, W. (1972). N=1:Diagnosis in Unique Cases, in Jacques, J.A. ed - Computer Diagnosis and Diagnostic Methods, Charles C. Thomas Pub., Springfield, IL.

Fleiss, J.L., Spitzer, R.L., Cohen, J. & Endicott, J. (1972). Three Computer Diagnosis Methods Compared, Archives of General Psychiatry, 27, 643-649.

Fryback, D.G. (1978). Bayes Theorem and Conditional Nonindependence of Data for Medical Diagnosis, Computers in Biomedical Research, 11, 423-434.

Gino, A., Pugh, W.M. & Ryman, D.H. (1990). MUMPS Based Integration of Disparate Computer-Assisted Medical Diagnosis Modules, Mumps User's Group Quarterly, 20, 68-74.

Goldman, L., Weinburg, M., Weisburg, R.N., Olsen, R. et al (1982). A Computer-Derived Protocol to Aid in the Diagnosis of Emergency Room Patients with Acute Chest Pain, New England Journal of Medicine, 17, 588-196.

Graham, W. (1977). Computer-aided Prediction of Gangrenous and Perforating Appendicitis, British Journal of Medicine, 34, 1375-1377.

Greenes, R.A., Tarabar, D.B., Krauss, M., Anderson, G., Wolnik, W.J. et al, (1989). Knowledge Management as a Decision Support Method: A Diagnostic Workup Strategy Application, Computers and Biolmedical Research, 22, 113-135.

Habbema, J.D. & Gelpke, G.J. (1987). A Computer Program for Selection of Variables in Diagnostic and Prognostic Problems, Computer Programs in Biomedicine, 13, 251-270.

Henderson, J.V., Moeller, G.M., Ryack, B.M. & Shumack, F., Adaptation of Computer-Assisted Diagnosis Program for use by Hospital Corpsmen Aboard Nuclear Submarines, Proceedings of 2nd Annual Symposium of Computer Application in Medical Care, Washington D.C., Nov. 5, 1978.

Hilden, J. (1984), Statistical Diagnosis Based on Conditional Independence Does not Require It, Computers in Biology & Medicine, 14, 429-435.

Hudson, D.L. & Cohen, M.E. (1987) The Impact of Reasoning with Uncertainty in a Medical Expert System, IEEE SCAMC, 43-48.

Hughes, D.A., Gose, E.E. & Roseman, D.C. (1990). Overcoming Deficiencies of the Rule Base Medical Expert Systems, Computer Methods & Programs in Biomedicine, 32, 63-72.

Kinney, E.L., Cortada, X., Galbut, D.L., Larsen, P.B., et al (1986). Use of the scientific literature to create a prototype expert system on unstable angina, Clinical Research, 34, 370.

Kinney, E.L., Brafman, D. & Wright, R.H. (1988). An Expert System on the Diagnosis of Ascites, Computers and Biomedical Research, 21, 169-173.

Kurzynski, M.W. (1987). Diagnosis of Acute Abdominal Pain Using a Three-Stage Classifier, Computers in Biology & Medicine, 17, 19-27.

Lachman, R. (1989). Expert Systems: A Cognitive Science Perspective, Behavior Research Methods,Instrumentation & Computers, 21, 195-204.

Ledley, R.S. & Lusted, L.B. (1959), Reasoning foundation of medical diagnosis: Symbolic Logic, Probability, and Value Theory aid in understanding of how physicians reason, Science, 130, 9-22.

Macartney, F.J. (1987) Diagnostic Logic, British Medical Journal, 295, 1325-1331.

Miller, R. A., Pople, H.E. & Myers, J.D. (1982). INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine, New England Journal of Medicine, 307, 468-476.

Miller, R., Masarie, F.E. & Myers, J.D. (1986). Quick Medical Reference (QMR) for Diagnostic Assistance, M.D. Computing, 3, 34-48.

Minasi, M. (1990). Bayes and Simple Expert Systems, AI Expert, 3, 13-15.

Nordyke, R.A., Kulikowski, C.A. & Kulikowski, C.A. (1971). A Comparison of Methods for the Automated Diagnosis of Thyroid Dysfunction, Computers and Biomedical Research, 4, 374-389.

Reggia, J.A. & Perricone, B.T. (1985). Answer Justification in Medical Decision Support Systems Based on Bayesian Classification, Computers in Biology & Medicine, 15 161-167.

Russek, E., Kromal, R.A. & Fisher, L.D. (1983). The Effect of Assuming Independence in Applying Bayes; Theorem to Risk Estimation and Classification in Diagnosis, Computers & Biomedical Research, 16, 537-552.

Schwartz, W.B., Petit, R.S. & Szolovits, P. (1987). A I in Medicine - Where do we stand?, New England Journal of Medicine, 316, 685-688.

Shortliffe, E.H., (1976). Computer-Based Medical Consultations: MYCIN, New York, Elsevier/North Holland.

Shortliffe, E.H. (1987). Computer Programs to Support Clinical Decision Making, JAMA, 258, 61-66.

Sneiderman, C.A. (1986), Computer Assisted Medical Diagnosis and Therapeutics: A Literature Review, Physicians & Computers, 16, 31-35.

Stetson, D.M., Eberhart, R.C., Dobbins, R.W., Pugh, W.M. & Gino, A., Structured Specification of a Computer Assisted Medical Diagnostic System, Proceedings Computer Assisted Medical Diagnostic System, Chapel Hill, North Carolina, June 12-14 1990.

Sturman, M.F. & Perez, M. (1989). Computer-Assisted Diagnosis of Acute Abdominal Pain, Comprehensive Therapy, 15, 26-35.

Tuhrim, S. & Reggia, J.A. (1986). Feasibility of Physician-developed Expert Systems, Medical Decision Making, 6, 23-26.

Weiss, S.M., Kulikowski, C.A., Amarel, S. & Safir (1978). A Model-Based Method for Computer-Aided Medical Decision Making, Artifical Intelligence, 11, 145-172.

Williams, B.T. ed., (1982). Computer Aids to Clinical Decisions, Vol. 2, CRC Press, Bocco Raton, Florida.

Wittokowski, K.M. (1990), Statistical Knowledge-Based Systems - Critical Remarks and Requirements for Approval, Computer Methods & Programs in Biomedicine, 33, 255-259.

Zagoria, R.J. & Reggia, J.A. (1983). Transferability of Medical Decision Support Systems Based on Bayesian Classification, Medical Decision Making, 3, 501-509.

Zoltie, N., Horrocks, J.C., de Dombal, F.T. (1977). Computer-assisted diagnosis of dyspepsia-report on transferability of a system, with emphasis on early diagnosis on gastric cancer, Methods Information Medicine, 16, 89-92.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE Feb 91	3. REPORT TYPE AND DATE COVERED Final

4. TITLE AND SUBTITLE
Computer Assisted Medical Diagnosis Problems, and Methods to Minimize their Effects

5. FUNDING NUMBERS
Program Element: 63706N
Work Unit Number:
M0095.005-6051

6. AUTHOR(S)
Ryman, David H.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No. 91-32

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The background and methods used in Computer Assisted Medical Diagnosis (CAMD) are dicussed and the problems and errors encountered during the development of CAMD systems are examined. Five specific methods and algorithms are discussed in terms of the problems previously reported in the literature, as well as, possible solutions and ways to minimize these problems. The use of multiple experts, adequate sampling sizes, multicenter samples with adequate and representative disease cases in the development of CAMD systems is recommended. Limitations of these systems can be determined from thorough system validation and testing of disease cases and sign/symptom data that were not used in the initial system development.

The utilization of sign and symptom disease complexes (indice, scales), and the use of multiple disease diagnostic methods to identify possible disease diagnosis is emphasized. Furthermore, CAMD systems should give standard definitions and interpretations for any signs, symptoms, lab results, diseases, or treatments suggested, and at a time the level of diagnostic expertise of the user should be considered. The development of the CAMD systems using a data base management approach that allow the integration of multiple methods is discussed as a valuable approach for implementing these systems.

14. SUBJECT TERMS
Computer Assisted Medical Diagnosis, diagnostic algorithms, Expert systems, Rule-based, Neural Network

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102